Détection d'anomalies des signaux AIS à partir de la fréquence instantanée

Steven COLLIN, Jean-Jacques SZKOLNIK, Abdel-Ouahab BOUDRAA, Delphine DARE, Cyril RAY

IRENav (EA3634), Ecole Navale/Arts-Métiers ParisTech, BCRM Brest, CC 600, 29240 Brest Cedex 9, France.
(steven.collin,jj.szkolnik,boudra,dare,ray)@ecole-navale.fr

Résumé – Cet article propose une approche pour détecter, de manière automatique, des anomalies de périodicité d'émission des messages AIS (Automatic Identification System) potentiellement révélatrices de comportements anormaux à partir de la seule analyse de la fréquence instantanée de l'enveloppe du signal. La fréquence instantanée est utilisée pour dater l'arrivée des messages et les classifier en fonction de l'identité de l'émetteur. Ensuite, un filtre de Kalman dont l'équation de dynamique modélise les récurrences potentielles en fonction de la classe du navire assure le suivi d'un type de message par émetteur. Si les récurrences subissent des variations inexpliquées, non prises en compte par le modèle, l'innovation augmente au delà d'un seuil fixé a priori et une alerte est remontée à un système expert de plus haut niveau.

Abstract – It is proposed to automatically detect anomalies of AIS (Automatic Identification System) messages repeatability potentially indicative of abnormal behavior from the analysis of the instantaneous frequency of the signal envelope. The instantaneous frequency is used to date the arrival of the messages and to classify them according to transmitter identity. Then, a Kalman filter whose dynamic equation models the potential recurrences according to the class of the ship, tracks one type of message per transmitter. If the recurrences present unexplained variations, not taken into account by the model, the innovation increases beyond a threshold and an alert is raised to a higher level expert system.

1 Introduction

Depuis quelques décennies, à l'instar de ce qui se passe dans d'autres secteurs, le domaine maritime subit de plein fouet une mutation numérique profonde et irréversible. Des systèmes automatisés d'échange d'informations ont vu le jour, l'AIS par exemple. Il permet, par liaison VHF, aux navires et systèmes de surveillance maritime de connaître un certain nombre d'informations à propos des navires présents dans une zone donnée, comme leur cinématique. Ce système présente des vulnérabilités importantes, plusieurs cas d'attaques ont déjà été recensés et des démonstrations de vulnérabilité réalisées [1],[2],[3],[4]. L'existence de failles de sécurité de l'AIS est maintenant actée et des travaux ont été réalisés sur le sujet. Ils consistent essentiellement à identifier les comportements "anormaux" par analyse du contenu des messages par rapport à des règles établies à partir de la fouille des données recueillies par les systèmes de gestion. La faisabilité d'utiliser le signal AIS en complément de la fouille de données pour détecter des anomalies comportementales est une thématique récente et peu développée. Les travaux de [5] en matière d'identification d'émetteur au travers de caractéristiques du niveau du signal AIS peuvent néanmoins être cités. Nous montrons dans ce travail comment la fréquence instantanée (FI), qui une information pertinente caractérisant le signal, permet pour dater l'arrivée de tous les messages, reconnaître les messages en provenance d'un émetteur et, accessoirement identifier le type de message. A notre connaissance, il n'existe pas de travaux dans la littérature relatant l'exploitation de la FI pour détecter de possibles falsifications du signal AIS.

2 Analyse du signal AIS

L'AIS émet sur deux fréquences VHF spécifiques, la modulation est de type GMSK (*Gaussian Minimum Shift Keying*) d'indice de modulation théorique 0.5 avec un débit de 9600 bauds. Les trames de 256 bits sont codées en NRZI (*Non Return to Zero Inverted*) avec un contrôle de redondance cyclique de degré 16. Une description de la trame AIS et des différents champs ainsi que le nombre de bits qui les composent figure dans [6]. L'accès au système de communication est de type partage temporel (2250 *time slot* par minute) suivant la méthode de l'accès multiple à répartition dans le temps auto-organisé (SOTDMA : *Self Organizing Time Division Access*).

2.1 Généralités

L'analyse du signal AIS s'effectue à partir des composantes en phase I et en quadrature Q de l'enveloppe complexe du signal. Les signaux en bande de base, échantillonnés à 250 kHz présentés ici proviennent d'une plateforme hybride décrite en cours d'article. La figure 1 donne un exemple, pour un message, de la valeur des voies I et Q à partir desquelles est calculée la phase. La figure 2 illustre les valeurs prises par la phase pour deux messages différents. La phase varie de manière continue au cours du temps en fonction des symboles transmis. Plus que la valeur de la phase à un instant donné ce sont ses variations au cours du temps qui portent l'information, d'où l'intérêt de la FI, define comme suit :

$$f(t) = \frac{1}{2\pi}\frac{d\Phi(t)}{dt} \tag{1}$$

Les travaux présentés s'inscrivent dans le cadre du projet ANR DéAIS (Détection de messages AIS falsifiés et analyse des risques associés).

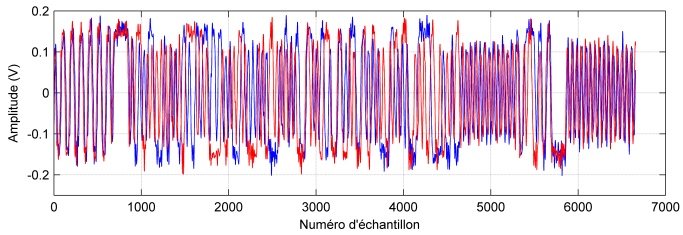

FIGURE 1 – Voie I (en bleu) et voie Q (en rouge) du signal AIS.

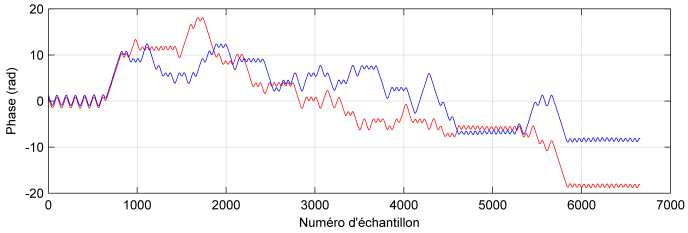

FIGURE 2 – Évolution de $\Phi(t)$ pour deux messages.

où $\Phi(t)$ est la phase du signal obtenue à partir des voies I et Q. La figure 3 illustre la FI.

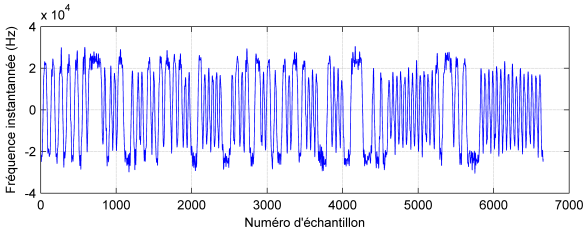

FIGURE 3 – FI $f(t)$ pour un message.

La détection et la datation des messages s'effectuent à l'aide d'un "motif" dit de référence (M_{Ref}), en l'occurrence la séquence des 24 bits de conditionnement commune à tous les messages, l'objectif consiste alors à la reconnaître dans le flux du signal reçu. Pour ce faire M_{Ref} est superposé à une portion du signal reçu de même longueur, un calcul de ressemblance est effectué puis M_{Ref} est translaté d'une période d'échantillonnage, et la même démarche est reproduite. La ressemblance est maximale lorsque M_{Ref} est superposé à une séquence de conditionnement. Pour l'identification du type de message, information codée sous 6 bits, la procédure est strictement identique, le M_{Ref} est alors fonction du type de message à reconnaître. Pour l'identification de l'émetteur d'un message, le M_{Ref} est déterminé à partir des 30 bits de l'indicatif (n°MMSI : *Maritime Mobile Service Identity*) de l'émetteur recherché, la procédure reste ensuite identique, tous les messages en provenance de l'émetteur auquel correspond le M_{Ref} sont ainsi identifiés. En répétant la procédure et en prenant en compte un nouveau M_{Ref} tant qu'il reste des messages non affectés à un émetteur il est possible de réaliser une partition des messages par émetteur. L'évaluation de la ressemblance est réalisée à l'aide d'une mesure de similarité.

2.2 Mesure de similarité

Pour la mesure de similarité, le choix s'est porté sur la fonction de corrélation, qui est plus robuste à l'effet d'échelle que la distance euclidienne [7]. Par ailleurs cette mesure est normalisée et par conséquent elle est indépendante de la valeur intrinsèque des échantillons. Un estimateur biaisé du coefficient de corrélation est donné par le coefficient de corrélation empirique, le nombre d'échantillons étant toutefois suffisant pour négliger ce biais.

2.2.1 Évaluation de la métrique à partir de la plateforme

L'évaluation de l'efficacité de la métrique à retrouver une séquence de référence dans un flux de données est réalisée à partir d'une la plate-forme hybride. Elle s'articule autour d'un ensemble comprenant deux équipements, un récepteur "RT820T Nooelec" et un émetteur-récepteur "HackRF" (*half-duplex*), et autour de la suite logicielle GNURadio (*free and open source*) dédiée à l'implémentation de radios logicielles.
Deux messages en provenance de deux émetteurs distincts (n° MMSI différents) sont générés à tour de rôle sur des *time slot* consécutifs. Le numéro MMSI de l'un des deux messages fait office de M_{Ref}. La figure 4 montre les résultats en utilisant la corrélation. Les maxima locaux de la métrique apparaissent clairement et sont séparés de la durée de deux *time slot* soit 53.3 ms. L'identification de l'émetteur recherché et des instants de réception du champ MMSI est faite de manière précise.

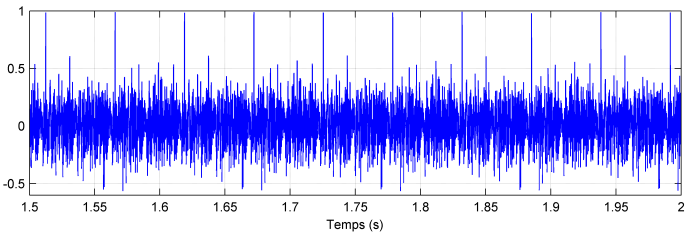

FIGURE 4 – Identification de l'émetteur par corrélation.

3 Exploitation de mesures réelles

Les signaux réels utilisés ont été enregistrés en 2015 en bande de base et échantillonnés à 25 kHz.

3.1 Détection des messages

L'évaluation des performances de détection est réalisée par comparaison du nombre de messages décodés lors de la campagne de mesure (fichiers AISlog) avec le nombre de messages détectés par la méthode basée sur le calcul de corrélation. Il apparaît que le calcul de corrélation détecte davantage de trames (entre 4% et 17 % de plus en fonction des enregistrements). La

Les travaux présentés s'inscrivent dans le cadre du projet ANR DéAIS (Détection de messages AIS falsifiés et analyse des risques associés).

figure 5 illustre ce phénomène, les messages détectés par corrélation coïncident davantage à la récurrence théorique du type du message que ceux figurant dans le fichier AISlog pour lesquels il existe des trous de détection potentiellement gênants.

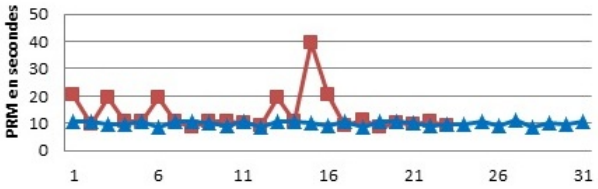

FIGURE 5 – Période de répétition des messages (fichier AISlog (rouge), algorithme (bleu).

3.2 Détection d'anomalies

La périodicité des temps d'arrivée des messages pour le système AIS obéit à des normes en fonction de la classe du navire et de sa cinématique et constitue par là un paramètre signant. La détection d'anomalies revient ainsi à repérer des altérations de la différence des temps d'arrivée des messages. La datation absolue de l'arrivée des messages n'amène pas d'information supplémentaire par rapport à la différence des temps d'arrivée qu'apporte la corrélation des numéros MMSI qui permet, en outre, d'obtenir une partition des messages par émetteur comme décrit au paragraphe 2.1.

3.2.1 Modèle d'état

Pour automatiser la détection d'anomalies de périodicité et afin de disposer d'une mesure de confiance de l'estimation de la périodicité, un filtre linéaire de Kalman est développé. Le système d'équations d'état est donné par (2), le vecteur d'état $\left[\begin{array}{cc} T_k & t_k \end{array}\right]'$ comprend deux variables d'état, la période d'émission T_k du message et le temps d'arrivée t_k du dernier message.

$$\left\{\begin{array}{l} \left[\begin{array}{c} T_{k+1} \\ t_{k+1} \end{array}\right] = \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array}\right]\left[\begin{array}{c} T_k \\ t_k \end{array}\right] + \left[\begin{array}{c} \beta_k \\ w_k \end{array}\right] \\ \\ z_k = \left[\begin{array}{cc} 0 & 1 \end{array}\right]\left[\begin{array}{c} T_k \\ t_k \end{array}\right] + \nu_k \end{array}\right. \quad (2)$$

avec :

- $\mathbf{Q} = \left[\begin{array}{cc} \sigma_\beta^2 & 0 \\ 0 & \sigma_w^2 \end{array}\right]$ la matrice de covariance du bruit d'état et $\mathbf{R} = \sigma_\nu^2$ la matrice de covariance du bruit de mesure.
- β_k est un bruit blanc centré gaussien permettant l'adaptation de la valeur de la période de répétition du message (PRM) à chaque itération du filtre. Son écart type est réglé de manière empirique lors de la phase de réglage du filtre.
- w_k représente un bruit blanc centré gaussien lié au matériel. Il permet de modéliser les incertitudes de modèle notamment la dérive de la source, ou le canal de propagation. Il est qualifié de *jitter* cumulatif, son écart type ne dépassant pas 1% de la périodicité des messages.
- ν_k est également un bruit blanc centré gaussien. Il représente le bruit de mesure de la représentation d'état et correspond à l'erreur de datation du $k^{\text{ème}}$ message.
- z_k est la mesure de l'instant d'arrivée du $k^{\text{ème}}$ message.

Les trois bruits sont supposés non corrélés entre eux.
Le filtre est basé sur un traitement séquentiel des temps d'arrivée des messages, il comprend deux phases distinctes, une phase de prédiction et une phase de mise à jour.

On pose $\mathbf{F} = \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array}\right]$ et $\mathbf{H} = \left[\begin{array}{cc} 0 & 1 \end{array}\right]$

Phase de prédiction :

$\hat{\mathbf{x}}_{k/k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1/k-1}$ (état prédit)
$\mathbf{P}_{k/k-1} = \mathbf{F}\mathbf{P}_{k-1/k-1}\mathbf{F} + \mathbf{Q}$ (covariance prédite)

Phase de mise à jour :

$\tilde{y}_k = z_k - \mathbf{H}\hat{\mathbf{x}}_{k/k-1}$ (innovation)
$\mathbf{S}_k = \mathbf{H}\mathbf{P}_{k/k-1}\mathbf{H}^T + \mathbf{R}$ (covariance de l'innovation)
$\mathbf{K}_k = \mathbf{P}_{k/k-1}\mathbf{H}^T\mathbf{S}_k^{-1}$ (gain de Kalman)
$\hat{\mathbf{x}}_{k/k} = \hat{\mathbf{x}}_{k/k-1} + \mathbf{K}_k\tilde{y}_k$ (état estimé)
$\mathbf{P}_{k/k} = (I - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k/k-1}$ (covariance erreur d'estimation)

Pour adapter la représentation d'état à la connaissance de l'expert, il est nécessaire de régler les variances des bruits d'état à partir de séquences d'apprentissage en observant la convergence, la stabilité et la consistance du filtre. A chaque message reçu, le filtre de Kalman met à jour les estimations du vecteur d'état et de l'erreur d'estimation par la prise en compte ou non du temps d'arrivée. Si $\tilde{y}_k > 3\sqrt{\mathbf{S}_k}$ la mesure n'est pas prise en compte et $\hat{\mathbf{x}}_{k/k} = \hat{\mathbf{x}}_{k/k-1}$ et $\mathbf{P}_{k/k} = \mathbf{P}_{k/k-1}$. Un tel comportement indique que la mesure n'obéit pas au modèle théorique des temps d'arrivée. Un nombre de rejets successifs supérieur à un seuil (5 en l'occurrence) implique l'arrêt du filtre en raison d'une probable anomalie de périodicité du message considéré.

3.2.2 Résultats obtenus à partir de données réelles

Compte tenu des données disponibles, seuls 76 navires possèdent un nombre de messages suffisants permettant l'exploitation de la périodicité. Ils sont classés en deux catégories. Ceux pour lesquels la PRM obéit aux spécifications du système AIS et qui engendrent un fonctionnement normal du filtre et ceux qui s'en écartent occasionnant un arrêt du filtre, en raison de messages manquants ou en surnombre. Pour les données utilisées la proportion totale d'anomalies est de 13%, compte tenu du seuil adopté.

Exemple de situation conforme : Cas d'un navire de classe A qui émet un message toutes les 10 secondes. La variabilité des PRM autour de la valeur théorique (Fig. 6) apparaît importante mais reste conforme aux spécifications (+/- 0,2 PRM).

Exemple de situation non conforme (messages manquants) : Dans l'exemple représenté figure 7, le nombre d'absences de messages consécutifs est supérieur au seuil fixé. Le modèle observé n'est donc plus conforme au modèle implémenté, ce qui provoque l'arrêt du filtre. L'émetteur est considéré comme disparu. La soudaine réapparition de l'émetteur (figure 7, courbe

Les travaux présentés s'inscrivent dans le cadre du projet ANR DéAIS (Détection de messages AIS falsifiés et analyse des risques associés).

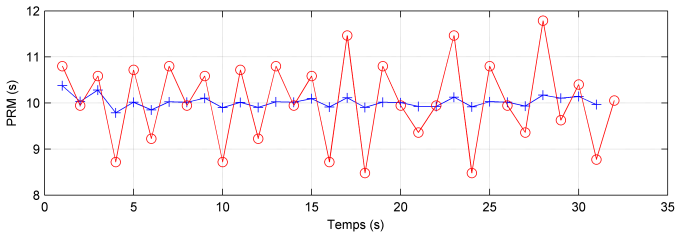

FIGURE 6 – PRM réelles (rouge), PRM estimées (bleu).

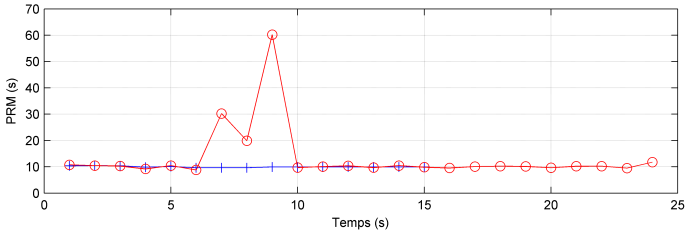

FIGURE 7 – PRM non conforme (messages manqués),PRM réelles (rouge), PRM estimées (bleu).

rouge) est traitée comme un nouvel émetteur et donne lieu à l'initialisation d'un nouveau filtre de Kalman. Il est ainsi possible de paramétrer le temps au bout duquel un message est considéré comme ayant disparu.

Exemple de situation non conforme (messages en surnombre) : Un autre type de situation non conforme survient lorsqu'au contraire du cas précédent, les messages sont plus nombreux qu'ils ne le devraient (parasites), figure 8. De la même manière que précédemment, la discordance du modèle observé avec le modèle implémenté provoque un arrêt du filtre et si la période de récurrence des messages redevient conforme aux spécifications du système, un nouveau filtre de Kalman est initialisé.

4 Conclusion

La fréquence instantanée de l'enveloppe complexe du signal AIS est une caractéristique intrinsèque au signal capable d'apporter des informations complémentaires au contenu des messages afin de détecter d'éventuelles falsifications du système. L'application d'une métrique basée sur le calcul séquentiel du coefficient de corrélation pour la détection des messages s'est révélée plus performante que celle basée sur le niveau d'énergie. La récurrence des messages constitue le seul critère d'analyse de falsification. Pour ce faire, les signaux réels sont classés par émetteurs et datés. Un filtre de Kalman est ensuite initialisé. Tant que la périodicité des messages correspond au modèle le filtre assure le suivi séquentiel de l'arrivée des messages en estimant la récurrence et l'erreur d'estimation. Cette dernière constitue une mesure de la confiance et est utilisée à un niveau de décision supérieur et fusionnée avec d'autres informations pour quantifier la probabilité d'une falsification. Lorsque qu'il manque des messages ou s'ils sont en surnombre, une analyse

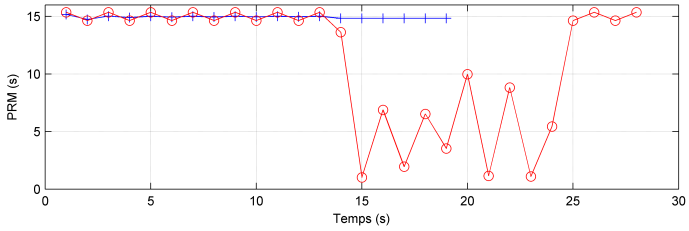

FIGURE 8 – PRM non conforme (messages parasites), PRM réelles (rouge), PRM estimées (bleu).

d'indicateurs de fonctionnement interne du filtre permet de déceler si le modèle implémenté correspond aux mesures. Au delà d'un seuil de rejet des mesures, une alerte est générée et vient enrichir le processus de détection de falsification. A court terme il est prévu d'évaluer une nouvelle métrique basée sur un calcul d'information mutuelle et de réaliser deux évolutions du filtre de Kalman. D'une part, l'introduction d'une fenêtre de validation des temps d'arrivée afin de ne garder que la mesure la plus vraisemblable (filtrage à association probabiliste), l'objectif étant d'augmenter la robustesse du filtre aux messages parasites (saturation). Et, d'autre part, transformer le filtre de Kalman affecté à chaque émetteur en un filtre à interaction de modèles multiples spécifique à chaque classe de navire (A ou B) lui conférant ainsi la capacité de suivre les évolutions de PRM liées aux changements de cinématique.

Références

[1] M. Balduzzi, A. Pasta and K. Wilhoit, "A Security evaluation of AIS automated identification system", *Proc. Annual Security Applications Conf.,* pp. 436-445, 2014.

[2] A. Harati-Mokhari, A. Wall, P. Brooks and J. Wang, "Automatic Identification System (AIS) : Data Reliability and Human Error Implications," *J. Navigation,* vol. 30, no. 3, pp. 373-389, 2007.

[3] F. Mazzarella, M. Vespe, D. Tarchi, G. Aulicino and A. Vollero, "AIS reception characterisation for AIS on/off anomaly detection,"*Int. Conf. Information Fusion,* pp. 1867-1873, 2016.

[4] F. Kastilieris, P. Braca and S. Coraluppi, "Detection of malicious AIS position spoofing by exploiting radar information," *Proc. Int. Conf. Information Fusion,* pp. 1196-1203, 2013

[5] E. Alincourt, C Ray, P-M Ricordel, D. Dare and A.O. Boudraa, "Methodology for AIS signature identification through magnitude and temporal characterization", OCEANS'16 MTS/IEEE, Shanghai, pp. 1-5, 2016.

[6] Union Internationale des Télécommunications, *ITU-R M.1371-5*,02/2014.

[7] J.Y. Stein, *Correlation,* in Digital Signal Processing : A Computer Science Perspective, John Wiley & Sons, Inc., 2000.

Les travaux présentés s'inscrivent dans le cadre du projet ANR DéAIS (Détection de messages AIS falsifiés et analyse des risques associés).